

CLAIMS

What is claimed is:

- 1 1. A method for mirroring data between a plurality of sites, comprising:
2 maintaining, at a first site of the plurality of sites, a record that identifies which
3 transactions that have been executed at the first site have had their redo
4 information replicated to the other sites of the plurality of sites;
5 determining a priority value associated with a transaction that is to be performed at
6 the first site, wherein the transaction specifies a modification to a data block;
7 if the priority value is a first value in a set of possible values, then committing the
8 transaction only after the record indicates that redo information associated
9 with the transaction has been replicated to the other sites of the plurality of
10 sites; and
11 if the priority value is a second value in said set of possible values, then committing
12 the transaction even though the record does not indicate that redo information
13 associated with the transaction has been replicated to the other sites of the
14 plurality of sites.
- 1 2. The method of Claim 1, wherein the first value indicates that the transaction should
2 not be lost if the first site becomes inoperable.
- 1 3. The method of Claim 1, wherein the second value indicates the transaction can be lost
2 if the first site becomes inoperable.
- 1 4. The method of Claim 1, further comprising the step of:
2 determining whether all other transactions that have committed before the transaction
3 has committed have had their respective redo information replicated to the

4 other sites of the plurality of sites by comparing a commit record associated
5 with the transaction to the record.

1 5. The method of Claim 1, wherein the record is a first record, and the method further
2 comprises the step of:

3 maintaining, at the first site, a second record that identifies which transactions that
4 have executed at the first site have had their redo information logged to
5 persistent storage at the first site.

1 6. The method of Claim 1, further comprising the step of:

2 if the priority value is the second value in the set of possible values, then committing
3 the transaction before the record indicates that the redo information generated
4 by the transaction has been replicated to the other sites of the plurality of sites.

1 7. The method of Claim 5, further comprising the step of:

2 if the priority value is the second value in the set of possible values, then committing
3 the transaction after the second record indicates that the redo information
4 generated by the transaction has been stored to persistent storage at the first
5 site.

1 8. The method of Claim 5, further comprising the step of:

2 determining which transactions that have executed at the first site have had their redo
3 information logged to persistent storage by comparing a commit record
4 associated with the transaction to the second record.

1 9. The method of Claim 1, further comprising the step of:

2 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
3 of the particular site after it is determined that all messages transmitted from
4 the particular site to each other site of the plurality of sites have been received
5 at their destination.

1 10. The method of Claim 1, further comprising the steps of:
2 at each site of the plurality of sites, determining if a data structure is to be replicated
3 to each other site of the plurality of sites; and
4 replicating the data structure to each other site of the plurality of sites unless it is
5 determined that the data structure is not to be replicated to each other site of
6 the plurality of sites.

1 11. A method for storing data, comprising:
2 at a first site in a plurality of sites, processing a transaction;
3 generating in volatile memory information that reflects the processed transaction; and
4 if said information has not been durably stored before either a data block associated
5 with the processed transaction is durably stored or the data block is transferred
6 to another site of the plurality of sites, then durably storing said information
7 before either the data block is durably stored or the data block is transferred to
8 another site of the plurality of sites.

1 12. The method of Claim 11, wherein the data block is a first data block, wherein the
2 transaction is a first transaction, the information is a first information, and the method
3 further comprises the steps of:
4 at the first site, processing a second transaction;

5 generating in volatile memory second information that reflects the processed second
6 transaction; and
7 if said first information and second information has not been durably stored before
8 either a second data block associated with the processed second transaction is
9 durably stored or the second data block is transferred to another site of the
10 plurality of sites, then durably storing using a batch process said first
11 information and said second information before either the second data block is
12 durably stored or the second data block is transferred to another site of the
13 plurality of sites.

1 13. The method of Claim 12, further comprising the step of:

2 determining whether the batch process has completed durably storing the first
3 information and the second information.

1 14. A method for mirroring data between a plurality of sites, comprising:

2 maintaining, at a first site of the plurality of sites, a record that identifies which
3 changes made to one or more data blocks stored at the first site have had
4 associated redo information replicated to the other sites of the plurality of
5 sites, wherein the first site implements a write-ahead logging scheme;
6 determining if the first site replicates, to the other sites of the plurality of sites, write
7 transactions that are executed at the first site in the order in which the write
8 transactions were issued; and
9 if the first site does not replicate, to the other sites of the plurality of sites, write
10 transactions that are executed at the first site in the order in which the write
11 transactions were issued, then durably storing a data block, in the one or more

12 data blocks, associated with a transaction only after the record indicates that
13 any write transactions that have updated the data block at the first site have
14 had their respective redo information replicated to the other sites of the
15 plurality of sites.

1 15. The method of Claim 14, wherein the record is a first record, and further comprising
2 the steps of:
3 maintaining, at the first site, a second record that identifies which changes made to
4 the one or more data blocks stored at the first site have had associated redo
5 information logged to persistent storage at the first site; and
6 if the first site does replicate, to the other sites of the plurality of sites, write
7 transactions that are executed at the first site in the order in which the write
8 transactions were issued, then durably storing the data block after the second
9 record indicates that any write transactions that have updated the data block at
10 the first site have had their respective redo information logged to persistent
11 storage at the first site.

1 16. The method of Claim 14, further comprising the step of:
2 releasing a lock associated with the data block after the first record indicates that redo
3 information associated with changes made to the data block has been
4 replicated to the other sites of the plurality of sites.

1 17. The method of Claim 15, wherein the first site replicates write transactions to the
2 other sites of the plurality of sites asynchronously to the completion of the write
3 transaction at the first site.

1 18. The method of Claim 14, further comprising the step of:

2 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
3 of the particular site after it is determined that all messages transmitted from
4 the particular site to each other site of the plurality of sites have been received
5 at their destination.

1 19. The method of Claim 14, further comprising the steps of:

2 at each site of the plurality of sites, determining if a data structure is to be replicated
3 to each other site of the plurality of sites; and
4 replicating the data structure to each other site of the plurality of sites unless it is
5 determined that the data structure is not to be replicated to each other site of
6 the plurality of sites.

1 20. A method for mirroring data between a plurality of sites, wherein the plurality of sites

2 includes a first site, comprising:
3 at the first site, durably storing a data block prior to durably storing redo information
4 about changes made to the data block; and
5 at the first site, durably storing the redo information after the changes have been
6 replicated to the other sites in the plurality of sites.

1 21. The method of Claim 20, wherein the data block is in a plurality of data blocks,

2 wherein changes made to the plurality of data blocks are performed by transactions
3 issued by a single process, and further comprising the step of:
4 determining if a set of transactions issued by the single process have completed,
5 wherein the set of transactions made the changes to the plurality of data
6 blocks.

- 1 22. The method of Claim 20, wherein the data block is in a plurality of data blocks,
2 wherein changes made to the plurality of data blocks are performed by transactions
3 issued by two or more processes, and further comprising the step of:
4 determining when the changes have been replicated to the other sites in the plurality
5 of sites.
- 1 23. The method of Claim 20, further comprising the step of:
2 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
3 of the particular site after it is determined that all messages transmitted from
4 the particular site to each other site of the plurality of sites have been received
5 at their destination.
- 1 24. The method of Claim 20, further comprising the steps of:
2 at each site of the plurality of sites, determining if a data structure is to be replicated
3 to each other site of the plurality of sites; and
4 replicating the data structure to each other site of the plurality of sites unless it is
5 determined that the data structure is not to be replicated to each other site of
6 the plurality of sites.
- 1 25. The method of Claim 1, wherein the record identifies which transactions that have
2 been executed at the first site have had their redo information replicated to the other
3 sites of the plurality of sites by identifying a portion of a redo log file, and wherein all
4 transactions reflected in the identified portion of the redo log file have been replicated
5 to the other sites of the plurality of sites.

1 26. The method of Claim 5, wherein the second record identifies which transactions that
2 have executed at the first site have had their redo information logged to persistent
3 storage at the first site by identifying a portion of a redo log file, and wherein all
4 transactions reflected in the identified portion of the redo log file have been logged to
5 persistent storage at the first site.

1 27. A machine-readable medium carrying one or more sequences of instructions for
2 mirroring data between a plurality of sites, wherein execution of the one or more
3 sequences of instructions by one or more processors causes the one or more
4 processors to perform the steps of:
5 maintaining, at a first site of the plurality of sites, a record that identifies which
6 transactions that have been executed at the first site have had their redo
7 information replicated to the other sites of the plurality of sites;
8 determining a priority value associated with a transaction that is to be performed at
9 the first site, wherein the transaction specifies a modification to a data block;
10 if the priority value is a first value in a set of possible values, then committing the
11 transaction only after the record indicates that redo information associated
12 with the transaction has been replicated to the other sites of the plurality of
13 sites; and
14 if the priority value is a second value in said set of possible values, then committing
15 the transaction even though the record does not indicate that redo information
16 associated with the transaction has been replicated to the other sites of the
17 plurality of sites.

1 28. The machine-readable medium of Claim 27, wherein the first value indicates that the
2 transaction should not be lost if the first site becomes inoperable.

1 29. The machine-readable medium of Claim 27, wherein the second value indicates the
2 transaction can be lost if the first site becomes inoperable.

1 30. The machine-readable medium of Claim 27, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 determining whether all other transactions that have committed before the transaction
5 has committed have had their respective redo information replicated to the
6 other sites of the plurality of sites by comparing a commit record associated
7 with the transaction to the record.

1 31. The machine-readable medium of Claim 27, wherein the record is a first record, and
2 wherein execution of the one or more sequences of instructions by the one or more
3 processors causes the one or more processors to further perform the step of:
4 maintaining, at the first site, a second record that identifies which transactions that
5 have executed at the first site have had their redo information logged to
6 persistent storage at the first site.

1 32. The machine-readable medium of Claim 27, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:

4 if the priority value is the second value in the set of possible values, then committing
5 the transaction before the record indicates that the redo information generated
6 by the transaction has been replicated to the other sites of the plurality of sites.

1 33. The machine-readable medium of Claim 31, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 if the priority value is the second value in the set of possible values, then committing
5 the transaction after the second record indicates that the redo information
6 generated by the transaction has been stored to persistent storage at the first
7 site.

1 34. The machine-readable medium of Claim 31, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 determining which transactions that have executed at the first site have had their redo
5 information logged to persistent storage by comparing a commit record
6 associated with the transaction to the second record.

1 35. The machine-readable medium of Claim 27, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
5 of the particular site after it is determined that all messages transmitted from
6 the particular site to each other site of the plurality of sites have been received
7 at their destination.

1 36. The machine-readable medium of Claim 27, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the steps of:
4 at each site of the plurality of sites, determining if a data structure is to be replicated
5 to each other site of the plurality of sites; and
6 replicating the data structure to each other site of the plurality of sites unless it is
7 determined that the data structure is not to be replicated to each other site of
8 the plurality of sites.

1 37. A machine-readable medium carrying one or more sequences of instructions for
2 storing data, wherein execution of the one or more sequences of instructions by one
3 or more processors causes the one or more processors to perform the steps of:
4 at a first site in a plurality of sites, processing a transaction;
5 generating in volatile memory information that reflects the processed transaction; and
6 if said information has not been durably stored before either a data block associated
7 with the processed transaction is durably stored or the data block is transferred
8 to another site of the plurality of sites, then durably storing said information
9 before either the data block is durably stored or the data block is transferred to
10 another site of the plurality of sites.

1 38. The machine-readable medium of Claim 37, wherein the data block is a first data
2 block, wherein the transaction is a first transaction, the information is a first
3 information, and wherein execution of the one or more sequences of instructions by
4 the one or more processors causes the one or more processors to further perform the
5 steps of:

6 at the first site, processing a second transaction;
7 generating in volatile memory second information that reflects the processed second
8 transaction; and
9 if said first information and second information has not been durably stored before
10 either a second data block associated with the processed second transaction is
11 durably stored or the second data block is transferred to another site of the
12 plurality of sites, then durably storing using a batch process said first
13 information and said second information before either the second data block is
14 durably stored or the second data block is transferred to another site of the
15 plurality of sites.

1 39. The machine-readable medium of Claim 38, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 determining whether the batch process has completed durably storing the first
5 information and the second information.

1 40. A machine-readable medium carrying one or more sequences of instructions for
2 mirroring data between a plurality of sites, wherein execution of the one or more
3 sequences of instructions by one or more processors causes the one or more
4 processors to perform the steps of:
5 maintaining, at a first site of the plurality of sites, a record that identifies which
6 changes made to one or more data blocks stored at the first site have had
7 associated redo information replicated to the other sites of the plurality of
8 sites, wherein the first site implements a write-ahead logging scheme;

9 determining if the first site replicates, to the other sites of the plurality of sites, write
10 transactions that are executed at the first site in the order in which the write
11 transactions were issued; and
12 if the first site does not replicate, to the other sites of the plurality of sites, write
13 transactions that are executed at the first site in the order in which the write
14 transactions were issued, then durably storing a data block, in the one or more
15 data blocks, associated with a transaction only after the record indicates that
16 any write transactions that have updated the data block at the first site have
17 had their respective redo information replicated to the other sites of the
18 plurality of sites.

1 41. The machine-readable medium of Claim 40, wherein the record is a first record, and
2 wherein execution of the one or more sequences of instructions by the one or more
3 processors causes the one or more processors to further perform the steps of:
4 maintaining, at the first site, a second record that identifies which changes made to
5 the one or more data blocks stored at the first site have had associated redo
6 information logged to persistent storage at the first site; and
7 if the first site does replicate , to the other sites of the plurality of sites, write
8 transactions that are executed at the first site in the order in which the write
9 transactions were issued, then durably storing the data block after the second
10 record indicates that any write transactions that have updated the data block at
11 the first site have had their respective redo information logged to persistent
12 storage at the first site.

1 42. The machine-readable medium of Claim 40, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 releasing a lock associated with the data block after the first record indicates that redo
5 information associated with changes made to the data block has been
6 replicated to the other sites of the plurality of sites.

1 43. The machine-readable medium of Claim 41, wherein the first site replicates write
2 transactions to the other sites of the plurality of sites asynchronously to the
3 completion of the write transaction at the first site.

1 44. The machine-readable medium of Claim 40, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
5 of the particular site after it is determined that all messages transmitted from
6 the particular site to each other site of the plurality of sites have been received
7 at their destination.

1 45. The machine-readable medium of Claim 40, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the steps of:
4 at each site of the plurality of sites, determining if a data structure is to be replicated
5 to each other site of the plurality of sites; and

6 replicating the data structure to each other site of the plurality of sites unless it is
7 determined that the data structure is not to be replicated to each other site of
8 the plurality of sites.

1 46. A machine-readable medium carrying one or more sequences of instructions for
2 mirroring data between a plurality of sites, wherein the plurality of sites includes a
3 first site, wherein execution of the one or more sequences of instructions by one or
4 more processors causes the one or more processors to perform the steps of:
5 at the first site, durably storing a data block prior to durably storing redo information
6 about changes made to the data block; and
7 at the first site, durably storing the redo information after the changes have been
8 replicated to the other sites in the plurality of sites.

1 47. The machine-readable medium of Claim 46, wherein the data block is in a plurality of
2 data blocks, wherein changes made to the plurality of data blocks are performed by
3 transactions issued by a single process, and wherein execution of the one or more
4 sequences of instructions by the one or more processors causes the one or more
5 processors to further perform the step of:
6 determining if a set of transactions issued by the single process have completed,
7 wherein the set of transactions made the changes to the plurality of data
8 blocks.

1 48. The machine-readable medium of Claim 46, wherein the data block is in a plurality of
2 data blocks, wherein changes made to the plurality of data blocks are performed by
3 transactions issued by two or more processes, and wherein execution of the one or

4 more sequences of instructions by the one or more processors causes the one or more
5 processors to further perform the step of:
6 determining when the changes have been replicated to the other sites in the plurality
7 of sites.

1 49. The machine-readable medium of Claim 46, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the step of:
4 if a particular site of the plurality of sites becomes inoperable, then initiating recovery
5 of the particular site after it is determined that all messages transmitted from
6 the particular site to each other site of the plurality of sites have been received
7 at their destination.

1 50. The machine-readable medium of Claim 46, wherein execution of the one or more
2 sequences of instructions by the one or more processors causes the one or more
3 processors to further perform the steps of:
4 at each site of the plurality of sites, determining if a data structure is to be replicated
5 to each other site of the plurality of sites; and
6 replicating the data structure to each other site of the plurality of sites unless it is
7 determined that the data structure is not to be replicated to each other site of
8 the plurality of sites.

1 51. The machine-readable medium of Claim 27, wherein the record identifies which
2 transactions that have been executed at the first site have had their redo information
3 replicated to the other sites of the plurality of sites by identifying a portion of a redo

4 log file, and wherein all transactions reflected in the identified portion of the redo log
5 file have been replicated to the other sites of the plurality of sites.

1 52. The machine-readable medium of Claim 31, wherein the second record identifies
2 which transactions that have executed at the first site have had their redo information
3 logged to persistent storage at the first site by identifying a portion of a redo log file,
4 and wherein all transactions reflected in the identified portion of the redo log file have
5 been logged to persistent storage at the first site.

1 53. The method of Claim 14, wherein the record identifies which changes are made to the
2 one or more data blocks stored at the first site have had associated redo information
3 replicated to the other sites of the plurality of sites by identifying a portion of a redo
4 log file, and wherein all changes in the identified portion of the redo log file have
5 been replicated to the other sites of the plurality of sites.

1 54. The method of Claim 15, wherein the second record identifies which changes are
2 made to the one or more data blocks stored at the first site have had associated redo
3 information logged to persistent storage by identifying a portion of a redo log file,
4 and wherein all changes in the identified portion of the redo log file have been logged
5 to persistent storage.

1 55. The machine-readable medium of Claim 40, wherein the record identifies which
2 changes are made to the one or more data blocks stored at the first site have had
3 associated redo information replicated to the other sites of the plurality of sites by
4 identifying a portion of a redo log file, and wherein all changes in the identified
5 portion of the redo log file have been replicated to the other sites of the plurality of
6 sites.

1 56. The machine-readable medium of Claim 41, wherein the second record identifies
2 which changes are made to the one or more data blocks stored at the first site have
3 had associated redo information logged to persistent storage by identifying a portion
4 of a redo log file, and wherein all changes in the identified portion of the redo log file
5 have been logged to persistent storage.